2009-03

Identifying the limits of training system effectiveness through taxonomies of human performance

Darken, R.P.

Taylor & Francis

Darken, R. P. "Identifying the limits of training system effectiveness through taxonomies of human performance." Theoretical Issues in Ergonomics Science 10.3 (2009): 231-243.

http://hdl.handle.net/10945/60924

Theoretical Issues in Ergonomics Science
Vol. 10, No. 3, May–June 2009, 231–243

Identifying the limits of training system effectiveness through taxonomies of human performance

R.P. Darken*

Naval Postgraduate School, Code CS/DR, Monterey, CA, USA

(*Received March 2006; final version received September 2007*)

Given a specific training problem, how much is known about choosing the best mix of technologies and methods that are both cost effective and that also meet given requirements? Beyond technical specifications of systems and components, a theoretically thorough method is still not used for identifying training requirements and how to best fit technology to meet those requirements. What is needed is a way to 'bound' training systems so that it can be clearly stated what they are and are not best used for. What are their limits? What are their strengths? The method described in this paper is based on a well-developed taxonomy of human performance. It is linked to conventional task analysis techniques to show how to identify what tasks a training system can be used for. An example is used to illustrate the method, but it has been applied to many training simulators. The process is meant to objectively link training requirements to technologies in a repeatable fashion such that training system builders, buyers and users can better understand the limits of their system.

Keywords: training; simulation; cognitive task analysis; human performance

1. The problem with training systems

Any reasonable study of the science behind training systems will indicate the importance of effectiveness evaluation. Building a system intended for training human performance does not imply that it will result in improved human performance. It seems obvious that in order to achieve this end, some form of training transfer study must be conducted to demonstrate that goals have been met, in terms of performance outcome, not technical attributes of the system. Yet, this is not done in many cases. Why not?

The most critical part of the problem is cultural, not technical or procedural. The training community still views training as a 'thing' not as an outcome. One typically leaps from training problem (a human performance shortfall of some type) directly to describing the system that is going to fix the problem without any real analysis of what the desired outcome is in terms of improved human performance. Training objectives are often an add-on that are described after the system is built rather than the driver for defining what the system will be. Unfortunately, many of the decision makers in training systems remain 'technology-centric' rather than 'human-centric'. One tends to measure effectiveness in

*Email: darken@nps.edu

ISSN 1463–922X print/ISSN 1464–536X online

DOI: 10.1080/14639220802151625
http://www.informaworld.com

terms of the features that the training system has rather than performance of the trainee on the transfer task. Even when training systems map into a training curriculum, there is often little supportable evidence that the outcome has been achieved. But before we get too critical of the development and evaluation process, we need to understand why this tendency exists.

It is well understood in the training community that training transfer studies are difficult to do. There are many reasons for this, among them:

- Access to an appropriate community of trainees is often problematic. Either they are difficult to reach (e.g. many military communities or first responders) or, in some cases, the trainee pool is very small (e.g. astronauts, government officials).
- Transfer studies usually require a high level of specialised expertise to design the study and to evaluate the data. A mix of subject matter expertise and experimental psychology is usually desirable, yet unavailable.
- It is often too risky to chance negative training on a treatment group. This is common in military settings. If the trainer turns out to be poor, what happens to the treatment group that used it? They probably have to recycle through the original training again, which is unacceptable with constrained budgets.
- The target trainee group does not actually measure anything on the transfer task, making a performance comparison purely subjective. A surprising number of cases have been found where no performance measurements were taken in the field. In these cases, traditional training is followed by a real-world event that is meant for experience only. So how would one know if one's trainer helped at all?
- Transfer studies are expensive to perform and can be extremely time and resource intensive.
- Transfer studies are susceptible to confounding variables and individual difference problems.
- Transfer studies are rarely part of the system design or integration cycle.

Another issue is that the high cost of development and fielding for a typical training system often causes programme managers to pool resources among multiple communities in order to justify the project. What can happen is that compromises are made by cooperating communities such that the resulting training system is highly functional but does not hit key perceptual cues or response mechanisms properly in order to facilitate effective learning for a particular training population. In order to be acceptable to multiple training communities, the system ends up sacrificing its potential effectiveness. The same thing can be said of training systems that is said of people – 'Jack of all trades, master of none'. An example from Naval Aviation would be mission rehearsal, where the same system is used by fixed and rotary wing pilots, even though mission profiles and perceptual cues are very different.

To further exacerbate the situation, the number of training systems requiring evaluation may be so large that it is simply not feasible to do a full transfer study on every system. While it might seem obvious that transfer studies should be the norm in the process for developing a training system, there are clearly some good reasons why they might be excluded. Is there anything that can be done to augment or replace a transfer study?

What the training community really needs are 'boundaries' for training systems. Given a training system, on which tasks is it effective and on which is it not effective? More specifically, on which task components is it effective? All training systems involve an abstraction of reality towards some training goal. Even in the best cases, there are many cues that a simulator will not replicate. Is it known what impact these omissions or simplifications have on training outcome? Most often, it is not. Further, training system developers are motivated to overstate, or at least make assumptions about, the effectiveness and usefulness of their training device. Walk an exhibition floor and there will be training systems to help make better decisions, get more rounds on target, obtain better situational awareness, etc. but rarely will there be any substance to these claims. In their defence, the ability of training system developers to conduct their own studies is likely highly limited. Further, it is not asserted that these trainers have no value, just that their value is poorly understood and that their bounds are unknown. As the recipient of such a device, how does the training manager decide how to use it?

There is an example of a process to bound a product's utility. The modelling and simulation community has developed a process for the verification, validation and accreditation (VV&A) of software models for simulations (Defense Modeling and Simulation Office 2005). These too are abstractions of reality and they are tested to ensure that the model's limitations are documented. The modelling and simulation VV&A process is certainly not without its flaws, but at least a process is in place that can be critiqued and improved and that is based on a scientific foundation. Such a process is needed for training systems. The solution needs to account for acquisition constraints as well. Specifically, programme managers care about cost, schedule and performance – not just performance. So investment in evaluating the bounds of a training system must have sufficient value to the overall programme.

Why, after many years of developing and studying training systems, are we still unable to reliably connect training technologies and learning strategies to training objectives? Given a training problem involving target acquisition from a helicopter, should the developer use a projection display system or a head-mounted display, or some other technology? What are the characteristics of the training problem that drive the developer towards one technology over another? Sometimes it is practical considerations such as cost or size, but if the deciding factor is the task itself, we are woefully weak on answers of substance backed with any empirical data. It is not that cost/benefit analyses are not done for training systems – they most certainly are. The problem is that the cost/benefit is not couched in the intended outcome of improved human performance. We tend to start with the technology part of the process, which is the fastest-changing part of the puzzle. As soon as we think we have learned something about displays, resolution or luminance goes up, and we are not sure what we know again. What would be desirable is a process that is tied to something more stable, independent of technology.

Training technologies are the fastest-changing part of the process, thus they are the worst choice by which to base the design of a training system. The elements of the task and the training requirements are more stable, but are still likely to change over time. The least likely component to change would be the trainee. While there is still much to learn, we know a lot about what a human being can and cannot do. Why not use the human's abilities as the basis for comparison and deciding on technologies for training? Furthermore, these abilities have been enumerated in well-developed taxonomies that will be useful (Fleishman and Quaintance 1984, Occupational Information Network 2006).

If it is impractical to conduct a training transfer study on every system that is built, then an improved process for developing training systems is needed, whereby we can know something substantive about the training system in terms of human performance before it is fielded. If a 'wish list' were to be built of what this process would need to do, it would include:

- *Objective and repeatable*: an assessment is of limited use if it is merely the evaluator's opinion or if another evaluator performing the same process comes up with a completely different answer. Subjective assessment of subject matter experts can be useful if focused on performance and not preference and if agreement between subject matter experts can be attained.
- *Human-centric*: one wants the assessment to focus on the trainee, not the technologies being used. This yields the greatest value since the technology will change but the trainee (and probably the task itself) will not.
- *Common language*: use a common basis for comparison so that as new systems are built and studied, more is known about how to build future systems.
- *Design recommendations*: help the system designer choose technologies and learning strategies based on similarities and differences to past analyses.
- *Bound the trainer*: help the designer articulate what the trainer is to be used for and what it is not to be used for based on known strengths and limitations. This applies equally well to part-task trainers where one needs to know which aspects of a larger task are covered and which are not.

The method described in this paper has been used to gain insight into a number of training systems in terms of the expected training outcome, the strengths and weaknesses of the system and guidelines for usage of the system (Cockayne and Darken 2004). Because so little is known about the mapping of sometimes subtle cues to training outcomes, one always tries to conduct transfer studies on every training system that is built. The process described here is an effective augmentation to a transfer study if a study can be performed. But if it cannot, it still yields insights into the strengths and limitations of the training system, thus informing training decisions.

2. Background

One of the most widely accepted models for training transfer is the training effectiveness ratio (Wickens and Hollands 2000). In this model, a measure is made of performance before and after training and the ratio is a quantitative measure of the change. It considers the time to train and how much performance may have improved over that time. There are two problems with this approach: (1) the method requires that a transfer study be done in every case, which is often not possible or practical; (2) the method tells the analyst how effective a training system might be but it does not have much to say as to why. We need to go a level deeper.

The learning outcome does indeed have a strong effect on the appropriate training apparatus. Magill (2001) describes motor skills as those requiring voluntary body or limb movement to achieve the goal. He further breaks these tasks down in terms of: (a) the size of musculature required (gross or fine); (b) the specificity of the task's beginning and ending point (discrete or continuous); (c) the stability of the environment (open or closed loop execution). Gentile's (2000) taxonomy breaks this down into the environment context (regulatory conditions and inertial variability) *vs* the function of the action (body orientation and object manipulation). Using either of these methods, one can clearly delineate between tasks such as hitting a baseball as opposed to hitting a golf ball. These tasks seem similar but the environmental characteristics (the fact that a golf ball is not moving when it is hit) make it possible to tease them apart in order to discover how motor learning takes place in each case. These techniques work very well for motor skills, but do not translate as well to cognitive skills such as decision making.

Bloom's taxonomy (Bloom 1956) was meant to classify forms of learning and it does so in a hierarchical paradigm whereby it is suggested that higher level learning cannot happen effectively until lower levels have been achieved. Bloom separated his taxonomy into cognitive, affective and psychomotor categories. In the cognitive domain, for example, the lowest level is knowledge where the learner acquires facts. This escalates to comprehension, application, analysis, synthesis and finally evaluation, where the learner might face a completely unique situation but can properly diagnose and respond to it effectively.

Merriënboer (1997) introduced the four-component instructional design (4C/ID) model for the development of training programmes for complex cognitive skills. The important elements of this model are that it differentiates between rule automation and schema acquisition. Rule automation is a linkage between stimulus and response through over-learning. Schema acquisition involves a deeper knowledge of the domain, a higher level of situational awareness and the ability to diagnose a situation. It typically requires some level of whole-task repetition, whereas rule automation can be accomplished with part-task trainers alone. To some extent, this sounds a lot like Bloom's taxonomy (Bloom 1956).

Both the rule automation and the schema acquisition components of the 4C/ID model have, as their input, some form of information flow. In rule automation, the learner perceives a specific stimulus (prerequisite information) and then responds accordingly. For schema acquisition, the knowledge is deeper but there is still a stream of information (supportive information) that must flow into whole-task practice for learning to take place. Sholl (1987) suggests that the schema also serves to guide perception. That is, one knows what information is missing and this is what one seeks to acquire, which then drives the decision-making process. In either case, if a simulator is used for part or whole-task practice, one needs to know what 'information' is flowing to the learner. Given that even the very best simulator that can possibly be created using modern technologies falls short of the full fidelity of the physical world, one needs to know what effect these differences might have on the training outcome.

Other alternative approaches to Bloom include Gagne's *The conditions of learning* (Gagne 1985). Gagne also layers learning, but the levels are purposely defined such that each has a different optimal type of instruction associated with it. His categories for learning include verbal information, intellectual skills, cognitive strategies, motor skills and attitudes. For example, to teach cognitive strategy, the learner must be able to develop solutions to problems and test them out.

Merrill (2002) advocates a problem-centred approach where the learner: (1) must activate prior experience as triggered by the problem; (2) demonstrate their skills as related to solving the problem; (3) apply their skills to the solution to the problem; (4) integrate their skills into real-world activities that solve the problem.

This topic will be returned to when the approach described in this paper is discussed, but what is required is a way for a customer to clearly state their training objectives with regard to the training system to be developed. There must be a 'language' by which these objectives are articulated. Bloom, Gagne, Merrill or some combination of these appear to be suitable mechanisms for this purpose.

The identical elements theory (Thorndike 1932) suggests that transfer will occur depending on there being some elements in common between the training task and transfer task. The greater the commonality, the greater the transfer. This has been significantly built upon in terms of identifying transferable (or generalisable) skills, but, even here, there is reliance on similarities between the training task and the transfer task. Enumerating all the possibilities, one sees that if a stimulus and its associated response do not match, training can be adversely affected.

If the stimulus is different but the response is the same ($S_{\cancel{m}}R_m$), then the trainee learns to trigger an action by the wrong cue. If it is the response that is mismatched ($S_mR_{\cancel{m}}$), then the trainee learns to do the wrong thing on the right cue. Also, of course, if both are mismatched ($S_{\cancel{m}}R_{\cancel{m}}$), no positive learning can take place. There is a close relationship between this idea and perceptual control theory, which describes a framework and approach to the analysis of human behaviour in order to determine what perceptual (not only visual) information is required for an operator to perform a task (Powers 1973, Riccio and McDonald 1998).

Yet it is known that there is much that cannot be done in simulations due to technical limitations. Does this mean that simulators are not useful for training? Of course they are, but how does one identify where potential pitfalls might be present? It could be argued that the first step is to identify where matches and mismatches occur. Then, where there is a technical (or other) reason why a mismatch must occur, it is imperative that this difference be made obvious to the trainee. For example, a recent training system built by the author's group was for Marine Corps Forward Observers. The resolution of the display was such that a very distant object could not be seen on the screen even when in reality it could be seen from that distance. These targets were highlighted to simplify the visual perception subtask of the trainee with the understanding that 'seeing the target' was not a training objective. Teaching the proper steps to neutralise the target was the purpose.

Schlager (1994) presents a 'task requirements matrix' and an 'instructional requirements matrix'. These are meant to identify where similarities are required and where they might not be required. However, the granularity of the analysis is too shallow for meaningful and repeatable results. The matrices map the operating environment, the participant, the activity and the requisite knowledge and skills directly to technologies such as displays, trackers and software. Schlager has the right idea, but not the right implementation. Any analysis he may have done is of little use today, more than 10 years later, because the technologies have all evolved since then. What is needed is a technique that links task elements at their lowest level directly to the human activity requirements demanded of those elements. At this level, it can be objectively stated how these human abilities (HAs) map to technologies. At this point, the requirements of the task can be compared to the capabilities of a training system – using the new approach described below – and, consequently, one can identify where the strengths and limitations are of the system, thus 'bounding' its capabilities.

3. Approach

The first step in the process is to separate inputs that need to come from the customer from those that are the domain of the training system designer (see Figure 1). (The term 'customer' is used to identify the person or organisation responsible for training and possibly funding the design and construction of the new training system, but not necessarily the actual person or population being trained.)

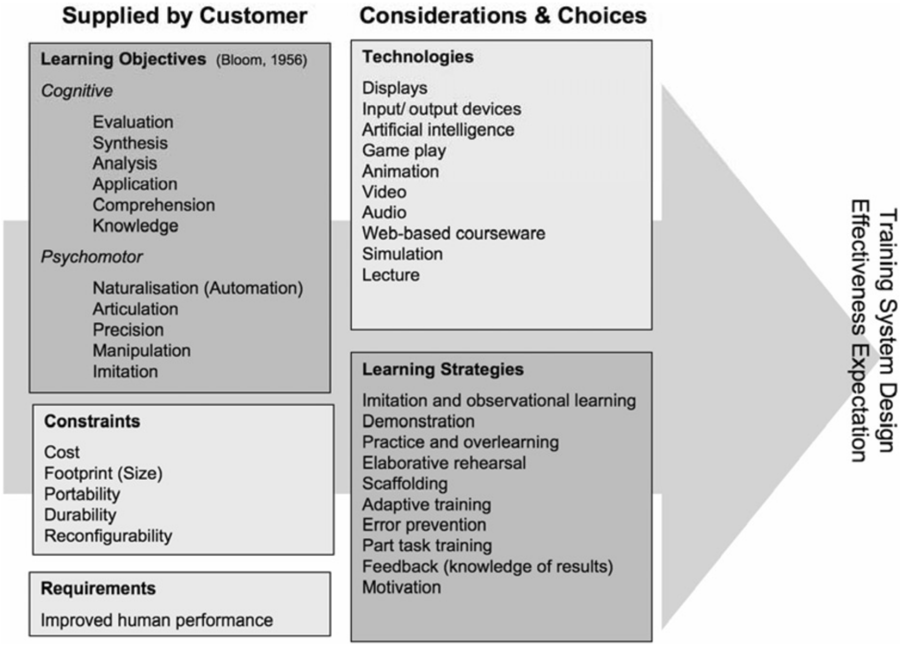

Figure 1. Components of the process of designing a training system. Some elements must be supplied by the customer, others need to be filled in by design decisions. Optimising this selection is the key to building effective training systems.

The customer may have constraints that affect the design process. Examples of these might be cost, footprint or portability requirement. These obviously limit design choices. Beyond this, the customer needs to identify the requirements – the tasks to be trained and the associated performance criteria. If there is an existing training system or process for this task, performance on that system might be used to set objectives for this new system. Is improved performance sought? Is faster convergence on 'trained' performance sought? In any case, there must be some metric that will be applied to trainees' post-training exposure so that the customer will know what training objectives are being met.

Finally, the customer needs to determine what the learning objectives are for this system. We are beginning to experiment with the use of Bloom's taxonomy (Bloom 1956) for this purpose as a way to designate what level of learning is the goal, but, as stated earlier, Gagne (1985) and Merrill (2002) offer other approaches that also appear to fit the paradigm very well. At the lowest level, one might want the trainee to acquire basic knowledge. At a higher level, one may want some level of diagnosis and synthesis requiring a higher level of understanding. This part of the process is not well developed yet so it is not discussed in detail. At a minimum, the customer should be able to set learning objectives for the system, whether or not Bloom's taxonomy is used to describe these. Ideally, learning objectives should inform design decisions or media selection (as suggested by Gagne).

More important than what customers should input to the process is what they should not input into the process. What the customer should not be defining is the solution. This is the domain of the training system designer. An additional component often overlooked is the need to also describe how the system should be used. This is referred to as learning strategies and has to do with feedback mechanisms, adaptation, scoring, tutoring, etc. This is equally important but beyond the scope of this paper. What will be focused on is the technology component and how selection of technologies is linked to the learning objectives and the tasks.

The technique described here is equally suitable for analysing existing training systems as it is for designing new ones. To analyse an existing trainer (this is referred to as a decomposition analysis), task elements and HAs are mapped to the individual components of the trainer. This tells us how the system should be bounded in terms of task elements that it can replicate and those that it cannot replicate. To design a new system (this is referred to as a composition analysis), a working knowledge of available technologies is needed. Here, task elements and HAs are mapped to the best (or most cost-effective) technologies that support the human performance needs of the task. In either case, the result is a bounding of the training device. This section will use an example decomposition analysis of the Indoor Simulated Marksmanship Trainer (ISMT) widely used by the Marine Corps for marksmanship skill development, shoot/don't shoot training and many other tasks involving the use of small arms. (This system is referred to as the ISMT, but the latest generation ISMT with new features is called ISMT-Enhanced (ISMT-E).)

The process begins with a thorough task analysis of the target training task. (CTA will be used herein to mean cognitive task analysis, but it is important that the analysis not be merely cognitive but must also identify perceptual cues and motor responses as required.)

It is important to carefully select the proper knowledge elicitation technique and knowledge representation scheme that yields easy access to the task hierarchy. Goals, operators, methods, selection criteria (GOMS) is used as a notation for task analyses (but only as a notation); however, there are other equally suitable methods (Card *et al.* 1983, John 1995). Maintaining the hierarchy is key, since it will be necessary to access the lowest level of the hierarchy in the next step, but the higher levels in the hierarchy help to identify which task elements are critical to the target training task and which may not be as central. In the example, a number of assumptions are made about marksmanship before conducting the analysis, such as daylight condition was assumed, it was assumed that the shooter was not moving and it was assumed that the weapon was loaded and ready and that it would not malfunction through the exercise. The GOMS analysis of the task was then completed.

The next step is to identify critical cues required in the execution of the task. The critical decision method (Klein *et al.* 1989) was used as it helps to identify not only the cues themselves but also the circumstances surrounding the use of those cues. These are important because they amplify the importance of these cues in executing the task. The simulator has much less room for error in critical cues than it might in some other aspect of the task.

If the CTA is considered to be the foundation of the process, one 'filters' it a number of times to either diminish or amplify the importance of certain cues or task components. Looking at the CTA through the training objectives 'filter' will tell us what task components are most important and which might be less important. The critical cues then tell us which cues, as they apply to specific key task elements of importance to training objectives, are most essential. For example, if wind effects are excluded from the marksmanship trainer as a training objective, then cues having to do with the shooter observing which way the wind might be blowing are less critical than they would be otherwise.

The next step requires the mapping of the lowest level of the CTA to what is referred to as an HA inventory. In other words, one wants to know what knowledge, skills and abilities are required of the human body in order to execute this task. The US Department of Labor has been interested in this question for many years as it is required to classify different jobs in terms of what is required to be able to perform that job. They (through a consortium) have developed a taxonomy of HAs that is decomposed into a skills inventory, a knowledge inventory and an abilities inventory (Occupational Information Network 2006). In the same way that the Department of Labor might describe a job in terms of these classifications, they are used to describe the training task.

Figure 2 shows an example of HA. Auditory attention is clearly defined and then it is to be evaluated for the given task in terms of its importance and then its level on a scaled spectrum. Using subject matter experts, the existence and importance of this ability for whatever task is being evaluated can be determined. The scale on the left simply asks how important this ability is to the task and the scale on the right estimates what level of involvement the ability has for the task.

In the marksmanship example, it was found that during the target identification component of the task, auditory attention was critical because the shooter might need to be able to hear spoken words or a weapon firing and that required high attention to what was being heard.

Once all the HAs that make up the task have been identified (all that are needed to fill out the CTA), the next step is to insert them in the appropriate places in the CTA so that it can be seen what HAs are needed for what task components. It is not to be assumed that the HA inventory will cover all aspects of the CTA that are required. This is why there is a facility for the expansion of the classification as required. In this analysis, it was discovered that while hearing sensitivity, auditory attention and speech recognition were all identified in the taxonomy, sound localisation was not. It was necessary to have the ability to locate a sound in the environment for the marksmanship task. For this, the HA inventory was extended. A definition for sound localisation was written and it was clearly differentiated

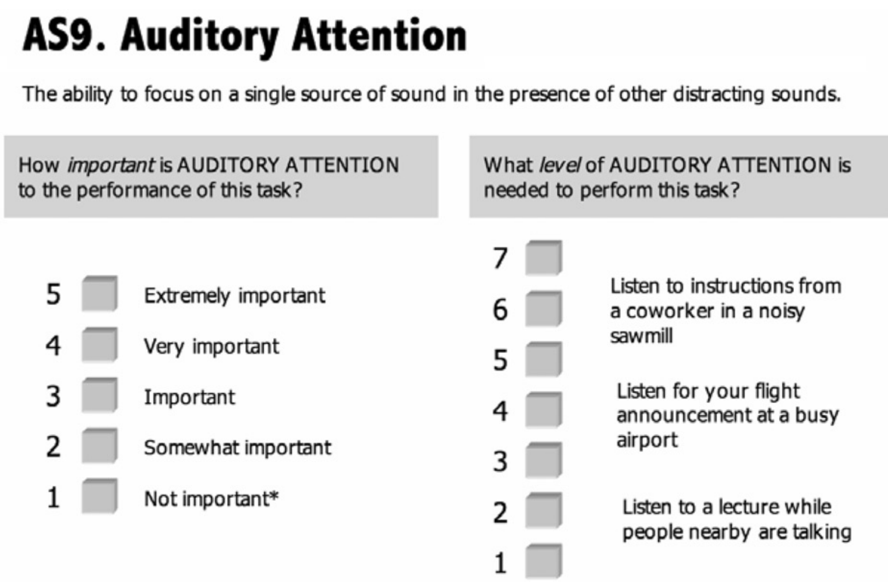

Figure 2. The auditory attention ability from O*NET (Occupational Information Network 2006).

from similar HAs. Several anchor points for the inventory were then identified so that it could be used and reused.

In the marksmanship example (see Figure 3), the original task analysis can be seen on the left with the corresponding HAs identified for the real-world task. This is a description of real marksmanship and the HAs required to execute that task. On the right is the same analysis, but on the ISMT training system. All the HAs that could not be supported by the ISMT system (marked with arrows) were identified. For example, the new HA, sound localisation (the ability to locate a sound in the environment) is required for the LISTEN-FOR-SOUNDS subtask, so it was tagged because the ISMT system only has stereo audio and cannot simulate spatial audio.

The HAs that are not supported by the ISMT apparatus are greyed out. Notice that one can easily see that there are components of the marksmanship task that are not well suited to ISMT. Since ISMT is not able to support cues to peripheral vision or depth perception, identifying hostile actions may be problematic. Observing weapons effects also may be difficult. The CTA suggests that to acquire a target, the marksman must scan the field of fire, which requires a wide field of view and depth perception, neither of which is supported by ISMT.

The full analysis of ISMT identified a number of strengths and weaknesses in the ISMT system. These are the 'bounds' for the ISMT trainer. It should not be used to train a shooter to scan a field of fire or for precise fire at a specific distance. It cannot handle wind effects. Most importantly, because it uses branching video and not computer-generated targets, it cannot simulate effects after fire so the shooter is unable to know what effects his fire may have had until the after-action review component of the training. On the other hand, the ISMT can train target identification within its limited field of fire. It excels in training proper body position, aiming, breathing and trigger pull. Its strengths are in the aspects most closely associated with the shooter, his body and the weapon. Weaknesses are concentrated in the surrounding environment.

Figure 3. The side-by-side view of the cognitive task analysis with human abilities for the real-world task (left) and the Indoor Simulated Marksmanship Trainer – Enhanced (ISMT-E) (right).

4. Conclusions

Revisiting the 'wish list' of a process for evaluating training systems, one can see if the process described here fills the requirements.

- *Objective and repeatable*. There is no denying that there is still some level of subjectivity in this process. Yet, by basing the evaluation on HAs that are well defined and placed in a structured taxonomy, the impact of subjectivity in the process is greatly limited.
- *Human-centric*. The taxonomy is based on HAs rather than on technology. Further, the taxonomy is extensible so that if a particular community (the military, for example) wanted to build HA definitions specific to its needs, this is possible.
- *Common language*. Using the HAs as that language, one can see how the evaluation of the ISMT (for example) will still be useful several years from now when display technologies and tracking are greatly improved because all one has to do is apply new technologies to the existing CTA with HAs. Marksmanship is likely to be very similar to what it is today.
- *Design recommendations*. Depending on what training objectives are of interest, the same CTA and HA lists can be used to help identify what technologies might

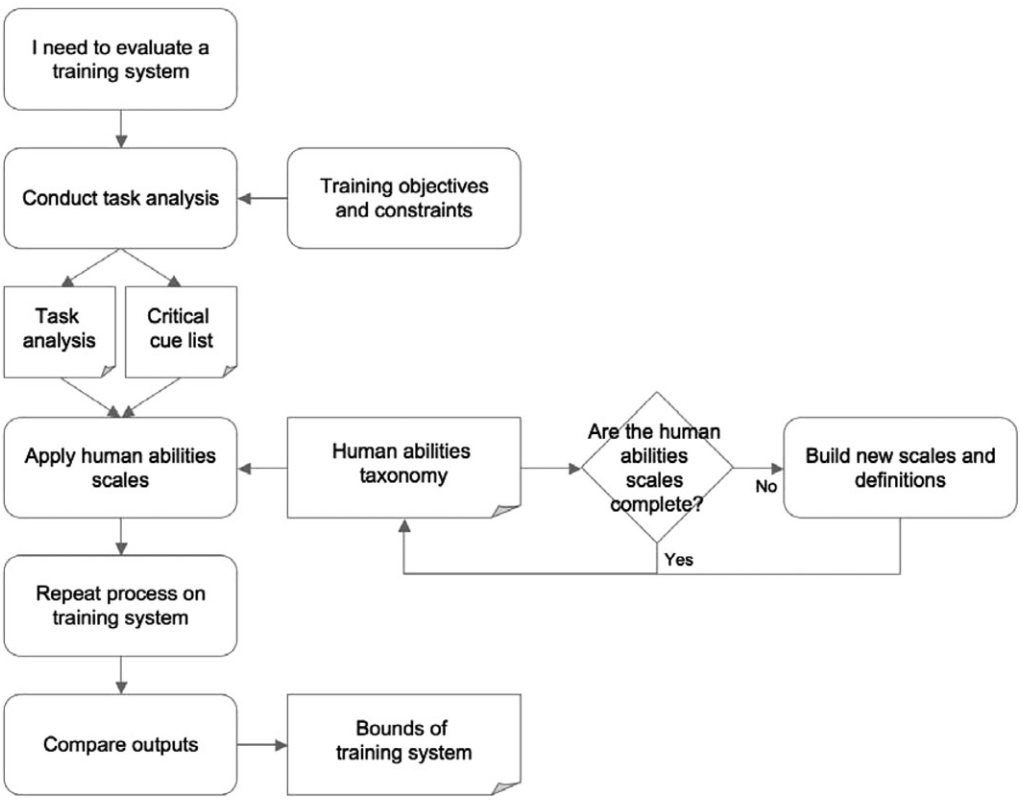

Figure 4. A flow chart of the process described in this paper.

best fit specific aspects of the task. If one wants ISMT to do something it is not equipped to do, the analysis can be used to show what needs to be improved to meet the new requirement.

- *Bound the trainer*. The list of limitations and strengths is the roadmap that bounds the ISMT trainer. It shows what aspects of the task are well supported and which are poorly supported.

The process still has several shortcomings. Recent focus has been on cognitive trainers where one can be allowed greater flexibility in some of the perceptual elements of the system. The literature on motor learning and cognitive learning are somewhat separate, but from the perspective of a user of a simulator for training, they are intimately related. A greater linkage of these literatures will be needed in order to bridge that gap. The process is also lacking in the application of learning strategies to training in addition to technologies at the stimulus/response level. Of the choices to be made in the development of a trainer, technologies are but one type of decision to be made. Feedback mechanisms, support for competitive learning and other learning strategies must also be applied. These have not yet been fitted into the process.

In summary, the process for identifying the limits of a training system through task decomposition is visually depicted in Figure 4. To evaluate a training system, the first step is to identify the constraints and the training objectives needed by the customer. These are inputs to the task analysis process that results in the CTA and associated critical cue lists. The next step is to apply the HA scales to the CTA and then repeat for the training system. If the HA taxonomy needs to be extended to fit the CTA, then define the needed HAs and insert into the taxonomy. Finally, compare the results of the real task to that of the training system and list the boundaries for the trainer. What should the trainer be used for? What tasks should be avoided? When is use limited to task familiarisation as opposed to performance measurement?

These questions need to be asked of training systems and the process for designing and analysing training systems must be able to answer them. The more one learns about the limits of training effectiveness of training systems, the more efficient one can be in what is built and how one uses training systems towards the development of highly skilled personnel.

References

Bloom, B.S., 1956. *Taxonomy of educational objectives, Handbook I: The cognitive domain*. New York: David McKay.

Card, S., Moran, T.P., and Newell, A., 1983. The GOMS model of manuscript editing. *In*: S. Card, T.P. Moran and A. Newell, eds. *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum, 139–147.

Cockayne, W.R. and Darken, R.P., 2004. The application of human ability requirements to virtual environment interface design and evaluation. *In*: D. Diaper and N. Stanton, eds. *The handbook of task analysis for human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum, 401–421.

Defense Modeling and Simulation Office, 2005. *VV&A recommended practices guide* [electronic version] [online]. Available from: http://vva.dmso.mil/[Accessed 22 August 2008].

Fleishman, E. and Quaintance, M., 1984. *Taxonomies of human performance: The description of human tasks*. Orlando, FL: Academic Press.

Gagne, R., 1985. *The conditions of learning*. 4th ed. New York: Holt, Rinehart & Winston.

Gentile, A.M., 2000. Skill acquisition: Action, movement, and neuromotor processes. *In*: J. Carr and R. Sheperd, eds. *Movement science: Foundations for physical therapy in rehabilitation*. 2nd ed. Gaithersburg, MD: Aspen, 111–186.

John, B.E., 1995. Why GOMS? *Interactions*, 2 (4), 80–89.

Klein, G.A., Calderwood, R., and MacGregor, D., 1989. Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, 19 (3), 462–472.

Magill, R.A., 2001. *Motor learning: Concepts and applications*. 6th ed. New York: McGraw-Hill.

Merrienboer, J.J.G.v., 1997. *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

Merrill, M.D., 2002. First principles of instruction. *Educational Technology Research and Development*, 50 (3), 43–59.

Occupational Information Network, 2006. *The O*NET content model* [online]. Available from: http://www.onetcenter.org [Accessed 22 August 2008].

Powers, W.T., 1973. *Behavior: The control of perception*. Chicago: Aldine.

Riccio, G.E. and McDonald, P.V., 1998. Technical Report NASA/TP 3703. *Multimodal perception and multicriterion control of nested systems: I. Coordination of postural control and vehicular control*. Houston, TX: Johnson Space Center.

Schlager, M.S., 1994. An approach for designing virtual environment training systems. *In*: *Proceedings of SIGCHI, Conference Companion*. New York: ACM Press, 317–318.

Sholl, M.J., 1987. Cognitive maps as orienting schemata. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13 (4), 615–628.

Thorndike, E., 1932. *The fundamentals of learning*. New York: Teachers' College Press.

Wickens, C.D. and Hollands, J.G., 2000. Memory, learning, and training. *In*: C.D. Wickens and J.G. Hollands, eds. *Engineering psychology and human performance*. 3rd ed. New York: Harper Collins, 241–284.

About the author

Rudolph Darken is the Director of the Defense Modeling and Simulation Institute (MOVES) and an Associate Professor of Computer Science at the Naval Postgraduate School in Monterey, California. He is the Chair of the MOVES Curriculum Committee and is also the Associate Director for Research for the Center for Homeland Defense and Security. His personal research has been primarily focused on human factors and training using virtual environments and computer-gaming media with emphasis on navigation and wayfinding in large-scale virtual worlds. He is a senior editor of *PRESENCE Journal*, the MIT Press journal of teleoperators and virtual environments. He received his BS in Computer Science Engineering from the University of Illinois at Chicago in 1990 and his MS and DSc degrees in Computer Science from The George Washington University in 1993 and 1995, respectively.